

Weekly Report

Lu Junhua

2016 年 12 月 26 日

Done

- Learned about what has been done in Yuquan. They didn't do much on data cleansing, e.g., removal of redundancy, semantically conflicts, etc. Their current strategy is simple. Besides, we discussed about potential ideas about detection.
- Brought back the data, I will store them on Spark, and start processing.
- Learned the progress of Yvquan team. Discuss potential visualization topic. Built a brief framework of course essay.

To do

- Process data, discuss more about the topic with Winson.
- Finish most parts of course essay.

论文

- InfoVis/TVCG 2016 *Many-to-Many Geographically-Embedded Flow Visualisation: An Evaluation* 文章主要是有一个很美观的设计, OD矩阵与起始地图结合, 对一些很小的细节做了优化, 比如矩阵旋转45度保持对称性, 扩展线之间距离等等以及一些有趣的交互. 此外还设计了相对完备的user study进行评估.
- *A C-LSTM Neural Network for Text Classification* 文本分类, 这里的C-LSTM其实思想很简单, 先用CNN抽出特征, 再在特征上做LSTM. 其分类效果(accuracy) 是好于其他的类似方法的, 除了比不上SVM. 但是SVM首先需要人力来进行特征划分, 不具有泛化能力, 但是本方法就不会.
- VAST 2016 *Magnostics: Image-based Search of Interesting Matrix Views for Guided Network Exploration* 探索矩阵中的特定模式, 利用30种现有的和自定义的feature descriptor, 计算特征. 对于这个设计特定的数据集, 用不同指标进行评估. 从用户场景说明使用特征描述子, 从大型网络数据中搜索具有某些特征的网络, 可以有效的引导用户分析探索这些数据(当然仍是这些相对简单的特征).
- SIGKDD 2002 *SPAM - Sequential Pattern Mining using A Bitmap Representation*对频繁时间序列的挖掘, 采用的思路有点类似与apriori. 但是时间序列的扩展有两种方法, 一种是s-扩展, 一种是i-扩展, 如何快速增加与扫描是一个问题. 利用bitmap来表示, 结合本身特性帮助使用者理解, 在可视化方面也能做一些文章.

- 其余看的是一些深度学习方面的文章, 针对体系结构课程. 有各种深度学习加速方法, 基于GPU的, 设计专用硬件的, 算法结合硬件进行改进的, 直接改进算法, 优化并行处理的. 不一一列举.